# Super-Resolution from Noisy Data

Emmanuel J. Candès* and Carlos Fernandez-Granda†

October 2012

**Abstract**

This paper studies the recovery of a superposition of point sources from noisy bandlimited data. In the fewest possible words, we only have information about the spectrum of an object in the low-frequency band $[-f_{\text{lo}}, f_{\text{lo}}]$ and seek to obtain a higher resolution estimate by extrapolating the spectrum up to a frequency $f_{\text{hi}} > f_{\text{lo}}$. We show that as long as the sources are separated by $2/f_{\text{lo}}$, solving a simple convex program produces a stable estimate in the sense that the approximation error between the higher-resolution reconstruction and the truth is proportional to the noise level times the square of the *super-resolution factor* (SRF) $f_{\text{hi}}/f_{\text{lo}}$.

**Keywords.** Deconvolution, stable signal recovery, sparsity, line spectra estimation, basis mismatch, super-resolution factor.

## 1 Introduction

It is often of great interest to study the fine details of a signal at a scale beyond the resolution provided by the available measurements. In a general sense, super-resolution techniques seek to recover high-resolution information from coarse scale measurements. There is a gigantic literature on this subject as researchers, for instance, always try to find ways of breaking the diffraction limit—a fundamental limit on the possible resolution—imposed by most imaging systems. Examples of applications include conventional optical imaging [18], astronomy [25], medical imaging [10], and microscopy [20]. In electronic imaging, photon shot noise limits the pixel size, making super-resolution techniques necessary to recover sub-pixel details [21, 23]. Among other fields demanding and developing super-resolution techniques, one could cite spectroscopy [11], radar [22], non-optical medical imaging [15] and geophysics [16].

In many of these applications, the signal we wish to super-resolve is a superposition of point sources; depending upon the situation, these may be celestial bodies in astronomy [19], molecules in fluorescence microscopy [19], or line spectra in speech analysis [14]. In the companion article [4], the authors studied the problem of deconvolving point sources from low-pass measurements. Whereas [4] focused mostly on the noiseless setting, in which one has perfect low-frequency information, this paper extends previous results by considering the noisy setting in which data are contaminated with noise, a situation which is unavoidable in practical situations. In a nutshell, [4] proves that with noiseless data, one can recover a superposition of point sources exactly, namely, with arbitrary high accuracy, by solving a simple convex program. This phenomenon holds as long as the spacing between the sources is on the order of the resolution limit. With noisy data now, it is of course no longer possible to achieve infinite precision. In fact, suppose the noise level and sensing resolution are fixed. Then one expects that it will become increasingly harder to recover the

*Departments of Mathematics and of Statistics, Stanford University, Stanford CA
†Department of Electrical Engineering, Stanford University, Stanford CA

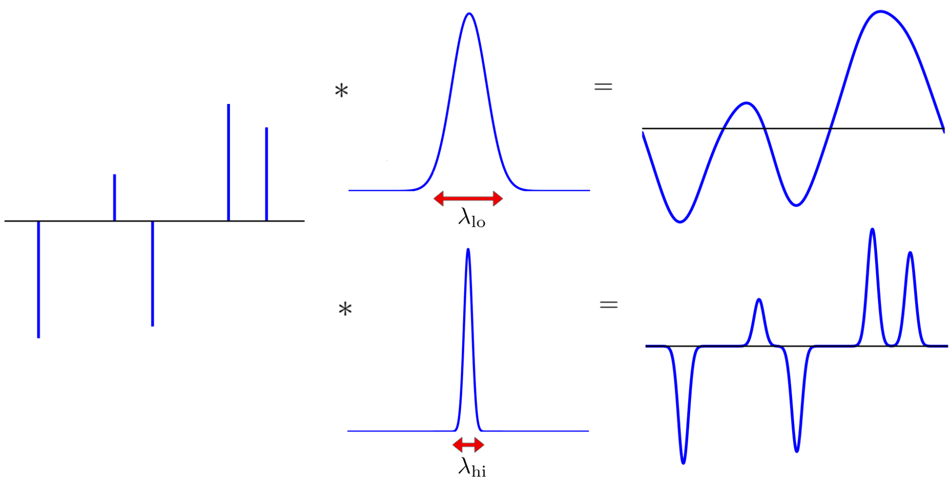



Figure 1: Sketch of the super-resolution factor (SRF). A signal (left) is measured at a low resolution by a convolution with a kernel (top middle) of width $\lambda_{\text{lo}}$ (top right). Super-resolution aims at approximating the outcome of a convolution with a much narrower kernel (bottom middle) of width $\lambda_{\text{hi}}$. Hence, the goal is to recover the bottom right curve.

fine details of the signal as the scale of these features become finer. The goal of this paper is to make this vague statement mathematically precise; we shall characterize the estimation error as a function of the noise level and of the resolution we seek to achieve. As we shall see next, increasing resolution essentially means filling-in parts of the missing spectrum.

### 1.1 The super-resolution problem

To formalize matters, we have observations about an object $x$ of the form

$$y(t) = (Q_{\text{lo}}x)(t) + z(t), \tag{1.1}$$

where $t$ is a continuous parameter (time, space, and so on) belonging to the $d$-dimensional cube $[0,1]^d$. Above, $z$ is a noise term which can either be stochastic or deterministic, and $Q_{\text{lo}}$ is a bandlimiting operator with a frequency cut-off equal to $f_{\text{lo}} = 1/\lambda_{\text{lo}}$. Here, $\lambda_{\text{lo}}$ is a positive parameter representing the finest scale at which $x$ is observed. To make this more precise, we take $Q_{\text{lo}}$ to be a low-pass filter of width $\lambda_{\text{lo}}$ as illustrated at the top of Figure 1; that is,

$$(Q_{\text{lo}}x)(t) = (K_{\text{lo}} * x)(t)$$

such that in the frequency domain the convolution equation becomes

$$(\widehat{Q_{\text{lo}}x})(f) = \widehat{K_{\text{lo}}}(f)\hat{x}(f), \quad f \in \mathbb{Z}^d,$$

where $\hat{x}(f) = \int e^{-i2\pi\langle f,t\rangle}x(\mathrm{d}t)$ is the usual Fourier transform. The low-pass kernel $\widehat{K_{\text{lo}}}(f)$ vanishes outside of the cell $[-f_{\text{lo}}, f_{\text{lo}}]^d$.

Our goal is to resolve the signal $x$ at a finer scale $\lambda_{\text{hi}} \ll \lambda_{\text{lo}}$. In other words, we would like to obtain a *high-resolution* estimate $x_{\text{est}}$ such that $Q_{\text{hi}}\,x_{\text{est}} \approx Q_{\text{hi}}\,x$, where $Q_{\text{hi}}$ is a bandlimiting operator with cut-off frequency $f_{\text{hi}} = 1/\lambda_{\text{hi}} > f_{\text{lo}}$. This is illustrated at the bottom of Figure 1, which shows the convolution between $K_{\text{hi}}$ and $x$. A different way to pose the problem is as follows: we have noisy data about the spectrum of an object of interest in the low-pass band $[-f_{\text{lo}}, f_{\text{lo}}]$, and would like to estimate the spectrum

in the possibly much wider band $[-f_{\text{hi}}, f_{\text{hi}}]$. We introduce the super-resolution factor (SRF) as:

$$\text{SRF} = \frac{f_{\text{hi}}}{f_{\text{lo}}} = \frac{\lambda_{\text{lo}}}{\lambda_{\text{hi}}}; \tag{1.2}$$

in words, we wish to double the resolution if the SRF is equal to 2, to quadruple it if the SRF equals four, and so on. Given the notorious ill-posedness of spectral extrapolation, a natural question is how small the error $K_{\text{hi}}(x_{\text{est}} - x)$ between the estimated and the true super-resolved signal at scale $\lambda_{\text{hi}}$ can be? In particular, how does it scale with both the noise level and the SRF? This paper addresses this important question.

## 1.2 Models and methods

As mentioned earlier, we are interested in superpositions of point sources modeled as

$$x = \sum_j a_j \delta_{t_j},$$

where $\{t_j\}$ are points from the interval $[0, 1]$, $\delta_\tau$ is a Dirac measure located at $\tau$, and the amplitudes $a_j$ may be complex valued. Although we focus on the one-dimensional case, our methods extend in a straightforward manner to the multidimensional case, as we shall make precise later on. We assume the model (1.1) in which $t \in [0, 1]$, which from now on we identify with the unit circle $\mathbb{T}$, and $z(t)$ is a bandlimited error term obeying

$$\|z\|_{L_1} = \int_{\mathbb{T}} |z(t)| \, \mathrm{d}t \leq \delta. \tag{1.3}$$

The measurement error $z$ is otherwise arbitrary and can be adversarial. For concreteness, we set $K_{\text{lo}}$ to be the periodic Dirichlet kernel

$$K_{\text{lo}}(t) = \sum_{k=-f_{\text{lo}}}^{f_{\text{lo}}} e^{i2\pi kt} = \frac{\sin(\pi(2f_{\text{lo}} + 1)t)}{\sin(\pi t)}. \tag{1.4}$$

By definition, for each $f \in \mathbb{Z}$, this kernel obeys $\widehat{K_{\text{lo}}}(f) = 1$ if $|f| \leq f_{\text{lo}}$ whereas $\widehat{K_{\text{lo}}}(f) = 0$ if $|f| > f_{\text{lo}}$. We emphasize, however, that our results hold for other low-pass filters. Indeed, our model (1.1) can be equivalently written in the frequency domain as $\hat{y}(f) = \hat{x}(f) + \hat{z}(f)$, $|f| \leq f_{\text{lo}}$. Hence, if the measurements are of the form $y = G_{\text{lo}} * x + z$ for some other low-pass kernel $G_{\text{lo}}$, then the model can be written as $\hat{y}(f) = \hat{x}(f) + \hat{z}(f)/\widehat{G_{\text{lo}}}(f)$, so that we have a very similar formulation. We omit the straightforward details.

To perform recovery, we propose solving

$$\min_{\tilde{x}} ||\tilde{x}||_{\text{TV}} \quad \text{subject to} \quad ||Q_{\text{lo}}\tilde{x} - y||_{L_1} \leq \delta. \tag{1.5}$$

Above, $||x||_{\text{TV}}$ is the total-variation norm of a measure (see Chapter 6 of [27] or Appendix A in [4]), which can be interpreted as the generalization of the $\ell_1$ norm to the real line. (If $x$ is a probability measure, then $||x||_{\text{TV}} = 1$.) This is not to be confused with the total variation of a function, a popular regularizer in signal processing and computer vision. Lastly, it is important to observe that the recovery algorithm is completely agnostic to the target resolution $\lambda_{\text{hi}}$, so our results hold simultaneously for any value of $\lambda_{\text{hi}} > \lambda_{\text{lo}}$.

## 1.3 Main result

Our objective is to approximate the signal up until a certain resolution determined by the width of the smoothing kernel $\lambda_{\text{hi}} > \lambda_{\text{lo}}$ used to compute the error. To fix ideas, we set

$$K_{\text{hi}}(t) = \frac{1}{f_{\text{hi}} + 1} \sum_{k=-f_{\text{hi}}}^{f_{\text{hi}}} (f_{\text{hi}} + 1 - |k|) \, e^{i2\pi kt} = \frac{1}{f_{\text{hi}} + 1} \left( \frac{\sin(\pi(f_{\text{hi}} + 1)t)}{\sin(\pi t)} \right)^2 \tag{1.6}$$

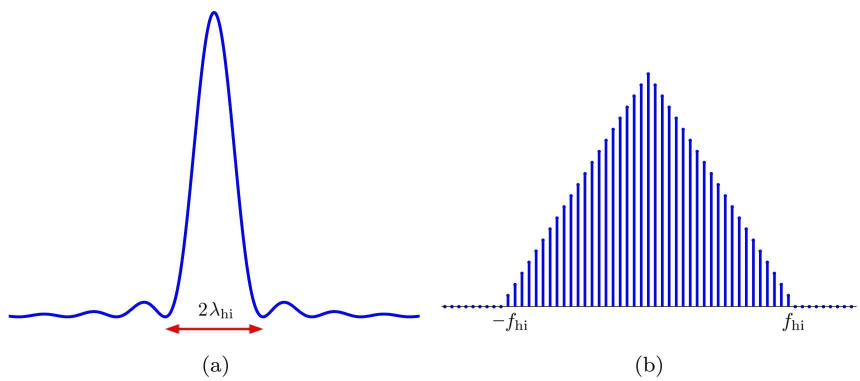



Figure 2: The Fejér kernel (1.6) (a) with half width about $\lambda_{\rm hi}$, and its Fourier series coefficients (b). The kernel is bandlimited since the Fourier coefficients vanish beyond the cut-off frequency $f_{\rm hi}$.

to be the Fejér kernel with cut-off frequency $f_{\rm hi} = 1/\lambda_{\rm hi}$. Figure 2 shows this kernel together with its spectrum.

As explained in Section 3.2 of [4], no matter what method is used to achieve super-resolution, it is necessary to introduce a condition about the support of the signal, which prevents the sources to be close to each other. Otherwise, the problem is easily shown to be hopelessly ill-posed by leveraging Slepian's work on prolate spheroidal sequences [31]. In this paper, we use the notion of minimum separation.

**Definition 1.1 (Minimum separation)** *For a family of points $T \subset \mathbb{T}$, the minimum separation is defined as the closest distance between any two elements from $T$,*

$$\Delta(T) = \inf_{(t,t') \in T \,:\, t \neq t'} |t - t'|.$$

Our model (1.3) asserts that we can achieve a low-resolution error obeying

$$||K_{\rm lo} * (x_{\rm est} - x)||_{L_1} \leq \delta,$$

but that we cannot do better as well. The main question is: how does this degrade when we substitute the low-resolution with the high-resolution kernel?

**Theorem 1.2** *Assume that the support $T$ of $x$ obeys the separation condition*

$$\Delta(T) \geq 2\lambda_{\rm lo}. \tag{1.7}$$

*Then under the noise model* (1.3)*, any solution $x_{\rm est}$ to problem* (1.5)[1] *obeys*

$$||K_{\rm hi} * (x_{\rm est} - x)||_{L_1} \leq C_0 \, \mathrm{SRF}^2 \, \delta,$$

*where $C_0$ is a positive numerical constant.*

Thus, minimizing the total-variation norm subject to data constraints yields a stable approximation of any superposition of Dirac measures obeying the minimum-separation condition. When $z = 0$, setting $\delta = 0$ and letting SRF$\to \infty$, this recovers the result in [4] which shows that $x_{\rm est} = x$, i.e. we achieve infinite precision. What is interesting here is the quadratic dependence of the estimation error in the super-resolution factor.

[1]To be precise, the theorem holds for any feasible point $\tilde{x}$ obeying $||\tilde{x}||_{\rm TV} \leq ||x||_{\rm TV}$; this set is not empty since it contains $x$.

It goes without saying that Theorem 1.2 can also be specialized to a stochastic noise model. Suppose we observe noisy samples of the spectrum

$$\eta(k) = \int_{\mathbb{T}} e^{-i2\pi kt}\, x(dt) + \epsilon_k, \quad k = -f_{\mathrm{lo}}, -f_{\mathrm{lo}} + 1, \ldots, f_{\mathrm{lo}}, \tag{1.8}$$

where $\epsilon_k$ is an iid sequence of complex-valued $\mathcal{N}(0, \sigma^2)$ variables (this means that the real and imaginary parts are independent $\mathcal{N}(0, \sigma^2)$ variables). This is equivalent to our model (1.1) with

$$z(t) = \sum_{k=-f_{\mathrm{lo}}}^{f_{\mathrm{lo}}} \epsilon_k e^{i2\pi kt}.$$

We have $\|z\|_{L_1} \leq \|z\|_{L_2}$ and $\|z\|_{L_2} = \|\epsilon\|_{\ell_2}$ by Parseval. Further, $||\epsilon||_2^2$ follows a $\chi^2$-distribution with $4f_{\mathrm{lo}} + 2$ degrees of freedom. As a result, a concentration inequality (see [17, Section 4]) yields

$$\mathbb{P}\left( ||\epsilon||_2 > (1+\gamma)\, \sigma \sqrt{4f_{\mathrm{lo}} + 2} \right) < e^{-2f_{\mathrm{lo}}\gamma^2},$$

for any positive $\gamma$. This gives the following corollary.

**Corollary 1.3** *Fix $\gamma > 0$. Under the stochastic noise model* (1.8)*, taking $\delta = (1+\gamma)\, \sigma \sqrt{4f_{\mathrm{lo}} + 2}$ yields*

$$||K_{\mathrm{hi}} * (x_{\mathrm{est}} - x)||_{L_1} \leq C_0\ (1+\gamma)\ \sqrt{4f_{\mathrm{lo}} + 2}\, \mathrm{SRF}^2\, \sigma. \tag{1.9}$$

*with probability at least $1 - e^{-2f_{\mathrm{lo}}\gamma^2}$.*

## 1.4 Extensions

*Other high-resolution kernels.* We work with the high-resolution Fejér kernel but our results would hold with just about any other symmetric kernel as long as the kernel obeys the properties (1.10) and (1.11) below as the proof only uses these simple estimates. The first reads

$$\int_{\mathbb{T}} |K_{\mathrm{hi}}(t)|\, \mathrm{d}t \leq C_0, \qquad \int_{\mathbb{T}} |K'_{\mathrm{hi}}(t)|\, \mathrm{d}t \leq C_1\, \lambda_{\mathrm{hi}}^{-1}, \qquad \sup |K''_{\mathrm{hi}}(t)| \leq C_2\, \lambda_{\mathrm{hi}}^{-3}, \tag{1.10}$$

where $C_0$, $C_1$ and $C_2$ are positive constants independent of $\lambda_{\mathrm{hi}}$. The second is that there exists a nonnegative and nonincreasing function $f : [0, 1/2] \to \mathbb{R}$ such that

$$|K''_{\mathrm{hi}}(t + \lambda_{\mathrm{hi}})| \leq f(t), \quad 0 \leq t \leq 1/2,$$

and

$$\int_0^{1/2} f(t)\mathrm{d}t \leq C_3\, \lambda_{\mathrm{hi}}^{-2}. \tag{1.11}$$

This is to make sure that (2.6) holds. (For the Fejér kernel, we can take $f$ to be quadratic in $[0, 1/2 - \lambda_{\mathrm{hi}}]$ and constant in $[1/2 - \lambda_{\mathrm{hi}}, 1/2]$.)

*Higher dimensions.* Our techniques can be applied to establish robustness guarantees for the recovery of point sources in higher dimensions. The only parts of the proof of Theorem 1.2 that do not generalize directly are Lemmas 2.4, 2.5 and 2.7. However, the methods used to prove these lemmas can be extended without much difficulty to multiple dimensions as described in Section C of the Appendix.

*Spectral line estimation.* Swapping time and frequency, Theorem 1.2 can be immediately applied to the estimation of spectral lines in which we observe

$$y(t) = \sum_j \alpha_j e^{i2\pi\omega_j t} + z(t), \quad t = 0, 1, \ldots, n-1,$$

where $z$ is a noise term. Here, our work implies that a nonparametric method based on convex optimization is capable of approximating the spectrum of a multitone signal with arbitrary frequencies, as long as these frequencies are sufficiently far apart, and furthermore that the reconstruction is stable. In this setting, the smoothed error can be interpreted as the recovery error windowed at a certain spectral resolution.

### 1.5 Related work

Since at least the work of Prony [24], parametric methods based on polynomial rooting have been a popular approach to the super-resolution of trains of spikes and, equivalently, of line spectra. These techniques are typically based on the eigendecomposition of a sample covariance matrix of the data [3, 26]. A statistical analysis of MUSIC [2, 28], a popular algorithm following this principle, can be found in [32] along with performance limits for any unbiased estimate based on a Cramér-Rao bound. More precise analysis has been carried out for models with a reduced number of parameters, yielding, for instance, a characterization of the trade-off between resolution and signal-to-noise ratio for the detection of two closely-spaced line spectra [30] or light sources [12,29]. In general, parametric techniques require prior knowledge of the model order and rely heavily on the assumption that the noise is white or at least has known spectrum (see Chapter 4 of [35]). An alternative approach that overcomes the latter drawback is to perform nonlinear least squares estimation of the model parameters [36]. Unfortunately, the resulting optimization problem has an extremely multimodal cost function, which makes it very sensitive to initialization [34]. Nonparametric methods based on convex programming do not require knowledge of the model order and are guaranteed to converge to a global optimum. However, previous theoretical work on the stability of this approach was limited to a discrete and finite-dimensional setting, where the support of the signal of interest is restricted to a finer uniform grid [4]. Other analyses of the super-resolution problem in the presence of noise also focus on signals supported on a grid [6, 31].

The total-variation norm is the continuous analog of the $\ell_1$ norm for finite dimensional vectors so that our recovery algorithm can be interpreted as finding the shortest linear combination—in an $\ell_1$ sense—of elements taken from a continuous and infinite dictionary. However, except for [4], previous stability results for sparse recovery in redundant dictionaries do not apply even if we discretize the dictionary; this is due to the high coherence between the elements. Moreover, working with a discrete dictionary can easily degrade the quality of the estimate [5] (see [33] for a related discussion concerning grid selection for spectral analysis). This observation has spurred the appearance of modified compressed-sensing techniques specifically tailored to the task of spectral estimation [7, 9, 13]. Proving stability guarantees for these methods under conditions on the support or the dynamic range of the signal is an interesting research direction.

## 2 Proof of Theorem 1.2

It is useful to first introduce various objects we shall need in the course of the proof. We let $T = \{t_j\}$ be the support of $x$ and define the disjoint subsets

$$\begin{aligned} S^{\lambda}_{\text{near}}(j) &:= \{t \;:\; |t - t_j| \le 0.16\lambda\}, \\ S^{\lambda}_{\text{far}} &:= \{t \;:\; |t - t_j| > 0.16\lambda,\ \forall t_j \in T\}; \end{aligned}$$

here, $\lambda \in \{\lambda_{\text{lo}}, \lambda_{\text{hi}}\}$, and $j$ ranges from 1 to $|T|$. We write the union of the sets $S^{\lambda}_{\text{near}}(j)$ as

$$S^{\lambda}_{\text{near}} := \cup_{j=1}^{|T|} S^{\lambda}_{\text{near}}(j)$$

and observe that the pair $(S^{\lambda}_{\text{near}}, S^{\lambda}_{\text{far}})$ forms a partition of $\mathbb{T}$. The value of the constant 0.16 is not important and chosen merely to simplify the argument. We denote the restriction of a measure $\mu$ with finite total variation on a set $S$ by $P_S\mu$ (note that in contrast we denote the low-pass projection in the frequency domain

by $Q_{\text{lo}}$). This restriction is well defined for the above sets, as one can take the Lebesgue decomposition of $\mu$ with respect to a positive $\sigma$-finite measure supported on any of them [27]. To keep some expressions in compact form, we set

$$
\begin{aligned}
I_{S^{\lambda}_{\text{near}}(j)}(\mu) &:= \frac{1}{\lambda_{\text{lo}}^2}\int_{S^{\lambda}_{\text{near}}(j)} (t-t_j)^2\,|\mu|\,(\mathrm{d}t)\,,\\
I_{S^{\lambda}_{\text{near}}}(\mu) &:= \sum_{t_j\in T} I_{S^{\lambda}_{\text{near}}(j)}(\mu)
\end{aligned}
$$

for any measure $\mu$ and $\lambda \in \{\lambda_{\text{lo}}, \lambda_{\text{hi}}\}$. Finally, we reserve the symbol $C$ to denote a numerical constant whose value may change at each occurrence.

Set $h = x - x_{\text{est}}$. The error obeys

$$||Q_{\text{lo}}h||_{L_1} \leq ||Q_{\text{lo}}x - y||_{L_1} + ||y - Q_{\text{lo}}x_{\text{est}}||_{L_1} \leq 2\delta,$$

and has bounded total-variation norm since $||h||_{\text{TV}} \leq ||x||_{\text{TV}} + ||x_{\text{est}}||_{\text{TV}} \leq 2\,||x||_{\text{TV}}$. Our aim is to bound the $L_1$ norm of the smoothed error $e := K_{\text{hi}} * h$,

$$||e||_{L_1} = \int_{\mathbb{T}}\left|\int_{\mathbb{T}} K_{\text{hi}}(t-\tau)\,h\,(\mathrm{d}\tau)\right|\mathrm{d}t.$$

We begin with a lemma bounding the total-variation norm of $h$ 'away' from $T$.

**Lemma 2.1** *Under the conditions of Theorem 1.2, there exist positive constants $C_a$ and $C_b$ such that*

$$
\begin{aligned}
\left|\left|P_{S^{\lambda_{\text{lo}}}_{\text{far}}}(h)\right|\right|_{\text{TV}} + I_{S^{\lambda_{\text{lo}}}_{\text{near}}}(h) &\leq C_a\,\delta,\\
\left|\left|P_{S^{\lambda_{\text{hi}}}_{\text{far}}}(h)\right|\right|_{\text{TV}} &\leq C_b\,\mathrm{SRF}^2\,\delta.
\end{aligned}
$$

This lemma is proved in Section 2.1 and relies on the existence of a low-frequency dual polynomial constructed in [4] to guarantee exact recovery in the noiseless setting.

To develop a bound about $\|e\|_{L_1}$, we begin by applying the triangle inequality to obtain

$$|e(t)| = \left|\int_{\mathbb{T}} K_{\text{hi}}(t-\tau)\,h\,(\mathrm{d}\tau)\right| \leq \left|\int_{S^{\lambda_{\text{hi}}}_{\text{far}}} K_{\text{hi}}(t-\tau)\,h\,(\mathrm{d}\tau)\right| + \left|\int_{S^{\lambda_{\text{hi}}}_{\text{near}}} K_{\text{hi}}(t-\tau)\,h\,(\mathrm{d}\tau)\right|. \tag{2.1}$$

By a corollary of the Radon-Nykodim Theorem (see Theorem 6.12 in [27]), it is possible to perform the polar decomposition $P_{S^{\lambda_{\text{hi}}}_{\text{far}}}(h)\,(\mathrm{d}\tau) = e^{i2\pi\theta(\tau)}\left|P_{S^{\lambda_{\text{hi}}}_{\text{far}}}(h)\right|(\mathrm{d}\tau)$ such that $\theta(\tau)$ is a real function and $\left|P_{S^{\lambda_{\text{hi}}}_{\text{far}}}(h)\right|$ is a positive measure. Then

$$
\begin{aligned}
\int_{\mathbb{T}}\left|\int_{S^{\lambda_{\text{hi}}}_{\text{far}}} K_{\text{hi}}(t-\tau)\,h\,(\mathrm{d}\tau)\right|\mathrm{d}t &\leq \int_{\mathbb{T}}\int_{S^{\lambda_{\text{hi}}}_{\text{far}}} |K_{\text{hi}}(t-\tau)|\left|P_{S^{\lambda_{\text{hi}}}_{\text{far}}}(h)\right|(\mathrm{d}\tau)\,\mathrm{d}t\\
&= \int_{S^{\lambda_{\text{hi}}}_{\text{far}}}\left(\int_{\mathbb{T}} |K_{\text{hi}}(t-\tau)|\,\mathrm{d}t\right)\left|P_{S^{\lambda_{\text{hi}}}_{\text{far}}}(h)\right|(\mathrm{d}\tau)\\
&\leq C_0\left|\left|P_{S^{\lambda_{\text{hi}}}_{\text{far}}}(h)\right|\right|_{\text{TV}}, \qquad (2.2)
\end{aligned}
$$

where we have applied Fubini's theorem and (1.10) (note that the total-variation norm of $\left|P_{S^{\lambda_{\text{hi}}}_{\text{far}}}(h)\right|$ is bounded by $2\,||x||_{\text{TV}} < \infty$).

In order to control the second term in the right-hand side of (2.1), we use a first-order approximation of the super-resolution kernel provided by the Taylor series expansion of $\psi(\tau) = K_{\text{hi}}(t-\tau)$ around $t_j$: for any $\tau$ such that $|\tau - t_j| \leq 0.16\lambda_{\text{hi}}$, we have

$$|K_{\text{hi}}(t-\tau) - K_{\text{hi}}(t-t_j) - K'_{\text{hi}}(t-t_j)(t_j-\tau)| \leq \sup_{u:|t-t_j-u|\leq 0.16\lambda_{\text{hi}}} \frac{1}{2}|K''_{\text{hi}}(u)|(\tau-t_j)^2.$$

Applying this together with the triangle inequality, and setting $t_j = 0$ without loss of generality, give

$$\begin{aligned}\int_{\mathbb{T}} \left| \int_{S^{\lambda_{\text{hi}}}_{\text{near}}(j)} K_{\text{hi}}(t-\tau)\, h(\mathrm{d}\tau) \right| \mathrm{d}t &\leq \int_{\mathbb{T}} \left| \int_{S^{\lambda_{\text{hi}}}_{\text{near}}(j)} K_{\text{hi}}(t)\, h(\mathrm{d}\tau) \right| \mathrm{d}t \\ &+ \int_{\mathbb{T}} \left| \int_{S^{\lambda_{\text{hi}}}_{\text{near}}(j)} K'_{\text{hi}}(t)\, \tau h(\mathrm{d}\tau) \right| \mathrm{d}t + \frac{1}{2}\int_{\mathbb{T}} \left| \int_{S^{\lambda_{\text{hi}}}_{\text{near}}(j)} \sup_{|t-u|\leq 0.16\lambda_{\text{hi}}} |K''_{\text{hi}}(u)|\, \tau^2 |h|(\mathrm{d}\tau) \right| \mathrm{d}t. \end{aligned} \tag{2.3}$$

(To be clear, we do not lose generality by setting $t_j = 0$ since the analysis is invariant by translation; in particular by a translation placing $t_j$ at the origin. To keep things as simple as possible, we shall make a frequent use of this argument.) We then combine Fubini's theorem with (1.10) to obtain

$$\int_{\mathbb{T}} \left| \int_{S^{\lambda_{\text{hi}}}_{\text{near}}(j)} K_{\text{hi}}(t)\, h(\mathrm{d}\tau) \right| \mathrm{d}t \leq \int_{\mathbb{T}} |K_{\text{hi}}(t)|\, \mathrm{d}t \left| \int_{S^{\lambda_{\text{hi}}}_{\text{near}}(j)} h(\mathrm{d}\tau) \right| \leq C_0 \left| \int_{S^{\lambda_{\text{hi}}}_{\text{near}}(j)} h(\mathrm{d}\tau) \right| \tag{2.4}$$

and

$$\int_{\mathbb{T}} \left| \int_{S^{\lambda_{\text{hi}}}_{\text{near}}(j)} K'_{\text{hi}}(t)\, \tau h(\mathrm{d}\tau) \right| \mathrm{d}t \leq \int_{\mathbb{T}} |K'_{\text{hi}}(t)|\, \mathrm{d}t \left| \int_{S^{\lambda_{\text{hi}}}_{\text{near}}(j)} \tau h(\mathrm{d}\tau) \right| \leq \frac{C_1}{\lambda_{\text{hi}}} \left| \int_{S^{\lambda_{\text{hi}}}_{\text{near}}(j)} \tau h(\mathrm{d}\tau) \right|. \tag{2.5}$$

Some simple calculations show that (1.10) and (1.11) imply

$$\int_{\mathbb{T}} \sup_{|t-u|\leq 0.16\lambda_{\text{hi}}} |K''_{\text{hi}}(u)|\, \mathrm{d}t \leq \frac{C_4}{\lambda^2_{\text{hi}}} \tag{2.6}$$

for a positive constant $C_4$. This together with Fubini's theorem yield

$$\begin{aligned}\int_{\mathbb{T}} \left| \int_{S^{\lambda_{\text{hi}}}_{\text{near}}(j)} |K''_{\text{hi}}(u)|\, \tau^2 |h|(\mathrm{d}\tau) \right| \mathrm{d}t &\leq \int_{\mathbb{T}} |K''_{\text{hi}}(t)|\, \mathrm{d}t \left| \int_{S^{\lambda_{\text{hi}}}_{\text{near}}(j)} \tau^2 |h|(\mathrm{d}\tau) \right| \\ &\leq C_4\, \mathrm{SRF}^2\, I_{S^{\lambda_{\text{hi}}}_{\text{near}}(j)}(h). \end{aligned} \tag{2.7}$$

In order to make use of these bounds, it is necessary to control the local action of the measure $h$ on a constant and a linear function. The following two lemmas are proved in Sections 2.2 and 2.3.

**Lemma 2.2** *Take $T$ as in Theorem 1.2 and any measure $h$ obeying $||Q_{\text{lo}} h||_{L_1} \leq 2\delta$. Then*

$$\sum_{t_j \in T} \left| \int_{S^{\lambda_{\text{hi}}}_{\text{near}}(j)} h(\mathrm{d}\tau) \right| \leq 2\delta + \left|\left| P_{S^{\lambda_{\text{hi}}}_{\text{far}}}(h) \right|\right|_{\text{TV}} + C\, I_{S^{\lambda_{\text{hi}}}_{\text{near}}}(h).$$

**Lemma 2.3** *Take $T$ as in Theorem 1.2 and any measure $h$ obeying $||Q_{\text{lo}} h||_{L_1} \leq 2\delta$. Then*

$$\sum_{t_j \in T} \left| \int_{S^{\lambda_{\text{hi}}}_{\text{near}}(j)} (\tau - t_j)\, h(\mathrm{d}\tau) \right| \leq C \left( \lambda_{\text{lo}}\, \delta + \lambda_{\text{lo}} \left|\left| P_{S^{\lambda_{\text{lo}}}_{\text{far}}}(h) \right|\right|_{\text{TV}} + \lambda_{\text{lo}}\, I_{S^{\lambda_{\text{lo}}}_{\text{near}}}(h) + \lambda_{\text{hi}}\, \mathrm{SRF}^2\, I_{S^{\lambda_{\text{lo}}}_{\text{near}}}(h) \right).$$

We may now conclude the proof of our main theorem. Indeed, the inequalities (2.2), (2.3), (2.4), (2.5) and (2.7) together with $I_{S^{\lambda_{\text{hi}}}_{\text{near}}}(h) \leq I_{S^{\lambda_{\text{lo}}}_{\text{near}}}(h)$ imply

$$||e||_{L_1} \leq C \left( \mathrm{SRF}\, \delta + \left|\left| P_{S^{\lambda_{\text{hi}}}_{\text{far}}}(h) \right|\right|_{\text{TV}} + \mathrm{SRF} \left|\left| P_{S^{\lambda_{\text{lo}}}_{\text{far}}}(h) \right|\right|_{\text{TV}} + \mathrm{SRF}^2\, I_{S^{\lambda_{\text{lo}}}_{\text{near}}}(h) \right) \leq C\, \mathrm{SRF}^2 \delta,$$

where the second inequality follows from Lemma 2.1.

## 2.1 Proof of Lemma 2.1

The proof relies on the existence of a certain low-frequency polynomial, and we first recall Proposition 2.1 and Lemma 2.5 from [4].

**Lemma 2.4** *Suppose $T$ obeys the separation condition (1.7) and take any $v \in \mathbb{C}^{|T|}$ with $|v_j| = 1$. Then there exists a low-frequency trigonometric polynomial*

$$q(t) = \sum_{k=-f_{\mathrm{lo}}}^{f_{\mathrm{lo}}} c_k e^{i2\pi kt}$$

*obeying the following properties:*

$$q(t_j) = v_j, \quad t_j \in T, \tag{2.8}$$

$$|q(t)| \leq 1 - \frac{C_a \left(t - t_j\right)^2}{\lambda_{\mathrm{lo}}^2}, \quad t \in S_{\mathrm{near}}^{\lambda_{\mathrm{lo}}}(j), \tag{2.9}$$

$$|q(t)| < 1 - C_b, \quad t \in S_{\mathrm{far}}^{\lambda_{\mathrm{lo}}}, \tag{2.10}$$

*with $0 < C_b \leq 0.16^2 C_a < 1$.*

Invoking a corollary of the Radon-Nykodim Theorem (see Theorem 6.12 in [27]), it is possible to perform a polar decomposition of $P_T h$,

$$P_T h = e^{i\phi(t)} \left|P_T h\right|,$$

such that $\phi(t)$ is a real function defined on $\mathbb{T}$. To prove Lemma 2.1, we work with $v_j = e^{-i\phi(t_j)}$. Since $q$ is low frequency,

$$\left|\int_{\mathbb{T}} q(t) \mathrm{d}h(t)\right| = \left|\int_{\mathbb{T}} q(t) Q_{\mathrm{lo}} h(t) \, \mathrm{d}t\right| \leq ||q||_{L_\infty} ||Q_{\mathrm{lo}} h||_{L_1} \leq 2\delta. \tag{2.11}$$

Next, since $q$ interpolates $e^{-i\phi(t)}$ on $T$,

$$\begin{aligned} ||P_T h||_{\mathrm{TV}} = \int_{\mathbb{T}} q(t) P_T h(\mathrm{d}t) &\leq \left|\int_{\mathbb{T}} q(t) h(\mathrm{d}t)\right| + \left|\int_{T^c} q(t) h(\mathrm{d}t)\right| \\ &\leq 2\delta + \sum_{j \in T} \left|\int_{S_{\mathrm{near}}^{\lambda_{\mathrm{lo}}}(j) \setminus \{t_j\}} q(t) h(\mathrm{d}t)\right| + \left|\int_{S_{\mathrm{far}}^{\lambda_{\mathrm{lo}}}} q(t) h(\mathrm{d}t)\right|. \end{aligned} \tag{2.12}$$

Applying (2.10) in Lemma 2.4 and Hölder's inequality, we obtain

$$\begin{aligned} \left|\int_{S_{\mathrm{far}}^{\lambda_{\mathrm{lo}}}} q(t) h(\mathrm{d}t)\right| &\leq \left|\left|P_{S_{\mathrm{far}}^{\lambda_{\mathrm{lo}}}} q\right|\right|_{L_\infty} \left|\left|P_{S_{\mathrm{far}}^{\lambda_{\mathrm{lo}}}}(h)\right|\right|_{\mathrm{TV}} \\ &\leq (1 - C_b) \left|\left|P_{S_{\mathrm{far}}^{\lambda_{\mathrm{lo}}}}(h)\right|\right|_{\mathrm{TV}}. \end{aligned} \tag{2.13}$$

Set $t_j = 0$ without loss of generality. The triangle inequality and (2.9) in Lemma 2.4 yield

$$\begin{aligned} \left|\int_{S_{\mathrm{near}}^{\lambda_{\mathrm{lo}}}(j) \setminus \{0\}} q(t) h(\mathrm{d}t)\right| &\leq \int_{S_{\mathrm{near}}^{\lambda_{\mathrm{lo}}}(j) \setminus \{0\}} |q(t)| \, |h| (\mathrm{d}t) \\ &\leq \int_{S_{\mathrm{near}}^{\lambda_{\mathrm{lo}}}(j) \setminus \{0\}} \left(1 - \frac{C_a t^2}{\lambda_{\mathrm{lo}}^2}\right) |h| (\mathrm{d}t) \\ &\leq \int_{S_{\mathrm{near}}^{\lambda_{\mathrm{lo}}}(j) \setminus \{0\}} |h| (\mathrm{d}t) - C_a I_{S_{\mathrm{near}}^{\lambda_{\mathrm{lo}}}(j)}(h). \end{aligned} \tag{2.14}$$

Combining (2.12), (2.13) and (2.14) gives

$$||P_T h||_{\mathrm{TV}} \leq 2\delta + ||P_{T^c} h||_{\mathrm{TV}} - C_b \left|\left| P_{S_{\mathrm{far}}^{\lambda_{\mathrm{lo}}}} (h) \right|\right|_{\mathrm{TV}} - C_a I_{S_{\mathrm{near}}^{\lambda_{\mathrm{lo}}}} (h) .$$

Observe that we can substitute $\lambda_{\mathrm{lo}}$ with $\lambda_{\mathrm{hi}}$ in (2.12) and (2.14) and obtain

$$||P_T h||_{\mathrm{TV}} \leq 2\delta + ||P_{T^c} h||_{\mathrm{TV}} - 0.16^2 \, C_a \, \mathrm{SRF}^{-2} \left|\left| P_{S_{\mathrm{far}}^{\lambda_{\mathrm{hi}}}} (h) \right|\right|_{\mathrm{TV}} - C_a I_{S_{\mathrm{near}}^{\lambda_{\mathrm{hi}}}} (h) .$$

This follows from using (2.9) instead of (2.10) to bound the magnitude of $q$ on $S_{\mathrm{far}}^{\lambda_{\mathrm{hi}}}$.

These inequalities can be interpreted as a generalization of the strong null-space property used to obtain stability guarantees for super-resolution on a discrete grid (see Lemma 3.1 in [4]). Combined with the fact that $\hat{x}$ has minimal total-variation norm among all feasible points, they yield

$$\begin{aligned} ||x||_{\mathrm{TV}} &\geq ||x + h||_{\mathrm{TV}} \\ &\geq ||x||_{\mathrm{TV}} - ||P_T h||_{\mathrm{TV}} + ||P_{T^c} h||_{\mathrm{TV}} \\ &\geq ||x||_{\mathrm{TV}} - 2\delta + C_b \left|\left| P_{S_{\mathrm{far}}^{\lambda_{\mathrm{lo}}}} (h) \right|\right|_{\mathrm{TV}} + C_a I_{S_{\mathrm{near}}^{\lambda_{\mathrm{lo}}}} (h) . \end{aligned}$$

As a result, we conclude that

$$C_b \left|\left| P_{S_{\mathrm{far}}^{\lambda_{\mathrm{lo}}}} (h) \right|\right|_{\mathrm{TV}} + C_a I_{S_{\mathrm{near}}^{\lambda_{\mathrm{lo}}}} (h) \leq 2\delta,$$

and by the same argument,

$$0.16^2 \, C_a \, \mathrm{SRF}^{-2} \left|\left| P_{S_{\mathrm{far}}^{\lambda_{\mathrm{hi}}}} (h) \right|\right|_{\mathrm{TV}} + C_a I_{S_{\mathrm{near}}^{\lambda_{\mathrm{hi}}}} (h) \leq 2\delta.$$

This finishes the proof.

## 2.2 Proof of Lemma 2.2

The proof also relies upon the low-frequency polynomial from Lemma 2.4 and the fact that $q(t)$ is close to $v_j$ when $t$ is near $t_j$. The intermediate result is proved in Section A of the Appendix.

**Lemma 2.5** *There is a polynomial $q$ satisfying the properties from Lemma 2.4 and, additionally,*

$$|q(t) - v_j| \leq \frac{C \left(t - t_j\right)^2}{\lambda_{\mathrm{lo}}^2}, \qquad \text{for all } t \in S_{\mathrm{near}}^{\lambda_{\mathrm{lo}}} (j) .$$

Consider the polar form

$$\int_{S_{\mathrm{near}}^{\lambda_{\mathrm{hi}}}(j)} h \, (\mathrm{d}\tau) = \left| \int_{S_{\mathrm{near}}^{\lambda_{\mathrm{hi}}}(j)} h \, (\mathrm{d}\tau) \right| e^{i\theta_j},$$

where $\theta_j \in [0, 2\pi)$. We set $v_j = e^{i\theta_j}$ in Lemma 2.4 and apply the triangular inequality to obtain

$$\begin{aligned} \left| \int_{S_{\mathrm{near}}^{\lambda_{\mathrm{hi}}}(j)} h \, (\mathrm{d}\tau) \right| &= \int_{S_{\mathrm{near}}^{\lambda_{\mathrm{hi}}}(j)} e^{-i\theta_j} h \, (\mathrm{d}\tau) \\ &\leq \int_{S_{\mathrm{near}}^{\lambda_{\mathrm{hi}}}(j)} q \left(\tau\right) h \, (\mathrm{d}\tau) + \left| \int_{S_{\mathrm{near}}^{\lambda_{\mathrm{hi}}}(j)} \left( q \left(\tau\right) - e^{-i\theta_j} \right) h \, (\mathrm{d}\tau) \right|, \end{aligned} \tag{2.15}$$

for all $t_j \in T$. By another application of the triangle inequality and (2.11)

$$\int_{S_{\mathrm{near}}^{\lambda_{\mathrm{hi}}}} q(\tau)\, h(\mathrm{d}\tau) \leq \left| \int_{\mathbb{T}} q(\tau)\, h(\mathrm{d}\tau) \right| + \left| \int_{S_{\mathrm{far}}^{\lambda_{\mathrm{hi}}}} q(\tau)\, h(\mathrm{d}\tau) \right| \leq 2\delta + \left|\left| P_{S_{\mathrm{far}}^{\lambda_{\mathrm{hi}}}}(h) \right|\right|_{\mathrm{TV}}. \tag{2.16}$$

To bound the remaining term in (2.15), we combine (2.9) in Lemma 2.4 and Hölder's inequality. With $t_j = 0$ (this is no loss of generality),

$$\begin{aligned} \left| \int_{S_{\mathrm{near}}^{\lambda_{\mathrm{hi}}}(j)} \left(q(t) - e^{-i\theta_j}\right) h(\mathrm{d}t) \right| &\leq \int_{S_{\mathrm{near}}^{\lambda_{\mathrm{hi}}}(j)} \left| q(t) - e^{-i\theta_j} \right| |h|(\mathrm{d}t) \\ &\leq \int_{S_{\mathrm{near}}^{\lambda_{\mathrm{hi}}}(j)} \frac{Ct^2}{\lambda_{\mathrm{lo}}^2} |h|(\mathrm{d}t) = C I_{S_{\mathrm{near}}^{\lambda_{\mathrm{hi}}}(j)}(h). \end{aligned}$$

It follows from this, (2.15) and (2.16) that

$$\left| \int_{S_{\mathrm{near}}^{\lambda_{\mathrm{hi}}}} h(\mathrm{d}\tau) \right| \leq 2\delta + \left|\left| P_{S_{\mathrm{far}}^{\lambda_{\mathrm{hi}}}}(h) \right|\right|_{\mathrm{TV}} + C I_{S_{\mathrm{near}}^{\lambda_{\mathrm{hi}}}}(h).$$

The proof is complete.

## 2.3 Proof of Lemma 2.3

We record a simple lemma.

**Lemma 2.6** *For any measure $\mu$ and with $t_j = 0$,*

$$\left| \int_{0.16\lambda_{\mathrm{hi}}}^{0.16\lambda_{\mathrm{lo}}} \tau \mu(\mathrm{d}\tau) \right| \leq 6.25\, \lambda_{\mathrm{hi}}\, \mathrm{SRF}^2\, I_{S_{\mathrm{near}}^{\lambda_{\mathrm{lo}}}(j)}(\mu).$$

**Proof** Note that in the interval $[0.16\lambda_{\mathrm{hi}}, 0.16\lambda_{\mathrm{lo}}]$, $t/0.16\,\lambda_{\mathrm{hi}} \geq 1$, whence

$$\left| \int_{0.16\lambda_{\mathrm{hi}}}^{0.16\lambda_{\mathrm{lo}}} \tau \mu(\mathrm{d}\tau) \right| \leq \int_{0.16\lambda_{\mathrm{hi}}}^{0.16\lambda_{\mathrm{lo}}} \tau\, |\mu|(\mathrm{d}\tau) \leq \int_{0.16\lambda_{\mathrm{hi}}}^{0.16\lambda_{\mathrm{lo}}} \frac{\tau^2}{0.16\lambda_{\mathrm{hi}}} |\mu|(\mathrm{d}\tau) \leq \frac{\lambda_{\mathrm{lo}}^2}{0.16\,\lambda_{\mathrm{hi}}} I_{S_{\mathrm{near}}^{\lambda_{\mathrm{lo}}}(j)}(\mu).$$

■

We now turn our attention to the proof of Lemma 2.3. By the triangle inequality,

$$\begin{aligned} &\sum_{t_j \in T} \left| \int_{S_{\mathrm{near}}^{\lambda_{\mathrm{hi}}}(j)} (\tau - t_j)\, h(\mathrm{d}\tau) \right| \leq \\ &\qquad \sum_{t_j \in T} \left| \int_{S_{\mathrm{near}}^{\lambda_{\mathrm{lo}}}(j)} (\tau - t_j)\, h(\mathrm{d}\tau) \right| + \sum_{t_j \in T} \left| \int_{0.16\lambda_{\mathrm{hi}} \leq |\tau - t_j| \leq 0.16\lambda_{\mathrm{lo}}} (\tau - t_j)\, h(\mathrm{d}\tau) \right|. \end{aligned} \tag{2.17}$$

The second term is bounded via Lemma 2.6. For the first, we use an argument very similar to the proof of Lemma 2.2. Here, we exploit the existence of a low-frequency polynomial that is almost linear in the vicinity of the elements of $T$. The result below is proved in Section B of the Appendix.

**Lemma 2.7** *Suppose $T$ obeys the separation condition (1.7) and take any $v \in \mathbb{C}^{|T|}$ with $|v_j| = 1$. Then there exists a low-frequency trigonometric polynomial*

$$q_1(t) = \sum_{k=-f_{\mathrm{lo}}}^{f_{\mathrm{lo}}} c_k e^{i2\pi kt}$$

*obeying*

$$|q_1(t) - v_j (t - t_j)| \leq \frac{C_a (t - t_j)^2}{\lambda_{\mathrm{lo}}}, \quad t \in S_{\mathrm{near}}^{\lambda_{\mathrm{lo}}}(j), \tag{2.18}$$

$$|q_1(t)| \leq C_b \lambda_{\mathrm{lo}}, \quad t \in S_{\mathrm{far}}^{\lambda_{\mathrm{lo}}}, \tag{2.19}$$

*for positive constants $C_a$, $C_b$.*

Consider the polar decomposition of

$$\int_{S_{\mathrm{near}}^{\lambda_{\mathrm{lo}}}(j)} (\tau - t_j)\, h(\mathrm{d}\tau) = \left| \int_{S_{\mathrm{near}}^{\lambda_{\mathrm{lo}}}(j)} (\tau - t_j)\, h(\mathrm{d}\tau) \right| e^{i\theta_j},$$

where $\theta_j \in [0, 2\pi)$, $t_j \in T$, and set $v_j = e^{i\theta_j}$ in Lemma 2.7. Again, suppose $t_j = 0$. Then

$$\begin{aligned}
\left| \int_{S_{\mathrm{near}}^{\lambda_{\mathrm{lo}}}(j)} \tau h(\mathrm{d}\tau) \right| &= \int_{S_{\mathrm{near}}^{\lambda_{\mathrm{lo}}}(j)} e^{-i\theta_j} \tau h(\mathrm{d}\tau) \\
&\leq \left| \int_{S_{\mathrm{near}}^{\lambda_{\mathrm{lo}}}(j)} \left( q_1(\tau) - e^{-i\theta_j} \tau \right) h(\mathrm{d}\tau) \right| + \int_{S_{\mathrm{near}}^{\lambda_{\mathrm{lo}}}(j)} q_1(\tau)\, h(\mathrm{d}\tau). \qquad (2.20)
\end{aligned}$$

The inequality (2.18) and Hölder's inequality allow to bound the first term in the right-hand side of (2.20),

$$\begin{aligned}
\left| \int_{S_{\mathrm{near}}^{\lambda_{\mathrm{lo}}}(j)} \left( q_1(\tau) - e^{-i\theta_j} \tau \right) h(\mathrm{d}\tau) \right| &\leq \int_{S_{\mathrm{near}}^{\lambda_{\mathrm{lo}}}(j)} \left| q_1(\tau) - e^{-i\theta_j} \tau \right| |h|(\mathrm{d}\tau) \\
&\leq \frac{C_a}{\lambda_{\mathrm{lo}}} \int_{S_{\mathrm{near}}^{\lambda_{\mathrm{lo}}}(j)} \tau^2 |h|(\mathrm{d}\tau) \\
&\leq C_a\, \lambda_{\mathrm{lo}}\, I_{S_{\mathrm{near}}^{\lambda_{\mathrm{lo}}}(j)}(h). \qquad (2.21)
\end{aligned}$$

Another application of the triangular inequality yields

$$\int_{S_{\mathrm{near}}^{\lambda_{\mathrm{lo}}}} q_1(\tau)\, h(\mathrm{d}\tau) \leq \left| \int_{\mathbb{T}} q_1(\tau)\, h(\mathrm{d}\tau) \right| + \int_{S_{\mathrm{far}}^{\lambda_{\mathrm{lo}}}} q_1(\tau)\, h(\mathrm{d}\tau). \tag{2.22}$$

We employ Hölder's inequality, (2.11), (2.18) and (2.19) to bound each of the terms in the right-hand side. First,

$$\left| \int_{\mathbb{T}} q_1(\tau)\, h(\mathrm{d}\tau) \right| \leq ||q_1||_{L_\infty} ||Q_{\mathrm{lo}} h||_{L_1} \leq C\, \lambda_{\mathrm{lo}}\, \delta. \tag{2.23}$$

Second,

$$\int_{S_{\mathrm{far}}^{\lambda_{\mathrm{lo}}}} q_1(\tau)\, h(\mathrm{d}\tau) \leq \left|\left| P_{S_{\mathrm{far}}^{\lambda_{\mathrm{lo}}}}(q_1) \right|\right|_{L_\infty} \left|\left| P_{S_{\mathrm{far}}^{\lambda_{\mathrm{lo}}}}(h) \right|\right|_{\mathrm{TV}} \leq C_b\, \lambda_{\mathrm{lo}} \left|\left| P_{S_{\mathrm{far}}^{\lambda_{\mathrm{lo}}}}(h) \right|\right|_{\mathrm{TV}}. \tag{2.24}$$

Combining (2.17) with these estimates gives

$$\sum_{t_j \in T} \left| \int_{S_{\mathrm{near}}^{\lambda_{\mathrm{hi}}}(j)} (\tau - t_j)\, h(\mathrm{d}\tau) \right| \leq C \left( \lambda_{\mathrm{lo}}\, \delta + \lambda_{\mathrm{lo}} \left|\left| P_{S_{\mathrm{far}}^{\lambda_{\mathrm{lo}}}}(h) \right|\right|_{\mathrm{TV}} + \lambda_{\mathrm{lo}}\, I_{S_{\mathrm{near}}^{\lambda_{\mathrm{lo}}}}(h) + \lambda_{\mathrm{hi}}\, \mathrm{SRF}^2\, I_{S_{\mathrm{near}}^{\lambda_{\mathrm{lo}}}}(h) \right),$$

as desired.

# 3 Discussion

We have shown that we could extrapolate the spectrum of a superposition of point sources by convex programming and that the extrapolation error scales quadratically with the super-resolution factor. This is a worst case analysis since the noise has bounded norm but is otherwise arbitrary. Natural extensions would include stability studies using other error metrics and noise models. For instance, an analysis tailored to a stochastic model might be able to sharpen Corollary 1.3 and be more precise in its findings. In a different direction, our techniques may be directly applicable to related problems. An example concerns the use of the total-variation norm for denoising line spectra [1]. Here, it would be interesting to see whether our methods allow to prove better denoising performance under a minimum-separation condition. Another example concerns the recovery of sparse signals from a random subset of their low-pass Fourier coefficients [37]. Here, it is likely that our work would yield stability guarantees from noisy low-frequency data.

On the algorithmic side, suppose we use the $L_2$ norm to constrain the feasible set,

$$\min_{\tilde{x}} \; ||\tilde{x}||_{\text{TV}} \quad \text{subject to} \quad \|Q_{\text{lo}}\tilde{x} - y\|_{L_2} \leq \delta. \tag{3.1}$$

Then the dual problem takes the form

$$\max_{u \in \mathbb{C}^n} \; \text{Re}\left[(F_{\text{lo}}\, y)^* \, u\right] - \delta \, ||u||_2 \quad \text{subject to} \quad ||F_{\text{lo}}^* \, u||_{L_\infty} \leq 1,$$

where $n = 2f_{\text{lo}} + 1$ and $F_{\text{lo}}$ denotes the linear operator that maps a function to its first $n := 2f_{\text{lo}} + 1$ Fourier coefficients as in (1.8) so that $Q_{\text{lo}} = F_{\text{lo}}^* F_{\text{lo}}$. The dual can be recast as the semidefinite program (SDP)

$$\max_{u} \; \text{Re}\left[(F_{\text{lo}}\, y)^* \, u\right] - \delta \, ||u||_2 \quad \text{subject to} \quad \begin{bmatrix} Q & u \\ u^* & 1 \end{bmatrix} \succeq 0, \quad \sum_{i=1}^{n-j} Q_{i,i+j} = \begin{cases} 1, & j = 0, \\ 0, & j = 1, 2, \ldots, n-1, \end{cases} \tag{3.2}$$

where $Q$ is an $n \times n$ Hermitian matrix, leveraging a corollary to Theorem 4.24 in [8] (see also [1, 4, 37]). In most cases, this allows to solve the primal problem with high accuracy.

**Lemma 3.1** *Let* $(x_{\text{est}}, u_{\text{est}})$ *be a primal-dual pair of solutions to* (3.1)–(3.2). *For any* $t \in \mathbb{T}$ *with* $x_{\text{est}}(t) \neq 0$,

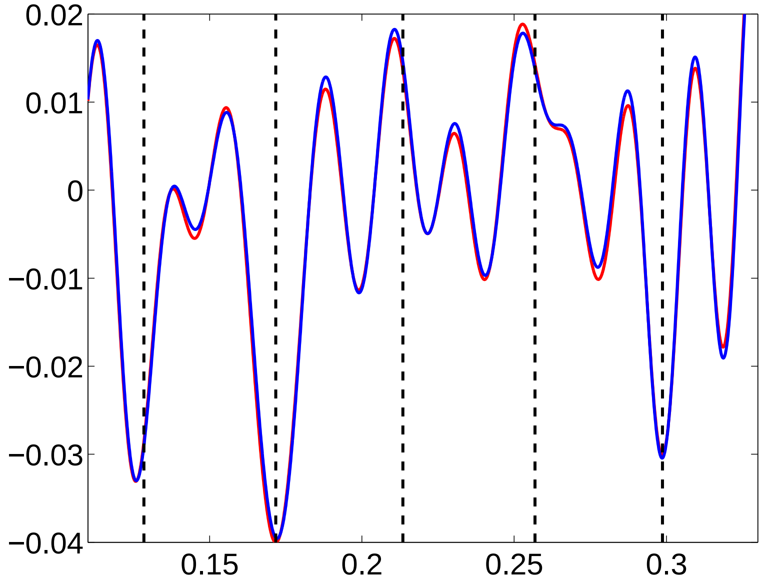
$$\left(F_{\text{lo}}^* \, u_{\text{est}}\right)(t) = \text{sign}\left(x_{\text{est}}(t)\right).$$

**Proof** First, we can assume that $y$ is low pass in the sense that $Q_{lo} y = y$. Since $x_{\text{est}}$ is feasible, $\|F_{\text{lo}}(y - x_{\text{est}})\|_{\ell_2} = \|y - Q_{\text{lo}} x_{\text{est}}\|_{L_2} \leq \delta$. Second, strong duality holds here. Hence, the Cauchy-Schwarz inequality gives

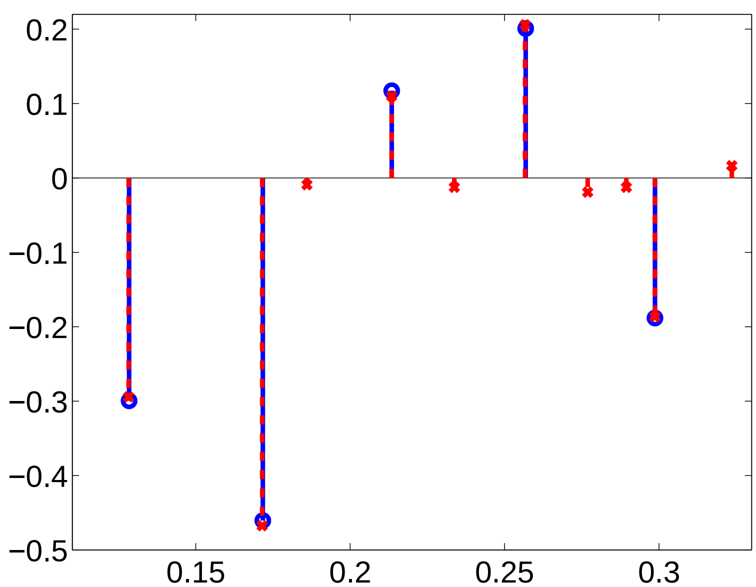
$$||x_{\text{est}}||_{\text{TV}} = \text{Re}\left[(F_{\text{lo}}\, y)^* \, u_{\text{est}}\right] - \delta \, ||u_{\text{est}}||_2 = \langle F_{\text{lo}} x_{\text{est}}, u_{\text{est}} \rangle + \langle F_{\text{lo}} y - F_{\text{lo}} x_{\text{est}}, u_{\text{est}} \rangle - \delta \, ||u_{\text{est}}||_2 \leq \langle x_{\text{est}}, F_{\text{lo}}^* u_{\text{est}} \rangle.$$

By Hölder's inequality and the constraint on $F_{\text{lo}}^* u_{\text{est}}$, $||x_{\text{est}}||_{\text{TV}} \geq \langle x_{\text{est}}, F_{\text{lo}}^* u_{\text{est}} \rangle$ so that equality holds. This is only possible if $F_{\text{lo}}^* u_{\text{est}}$ equals the sign of $x_{\text{est}}$ at every point where $x_{\text{est}}$ is nonzero. ■

This result implies that it is usually possible to determine the support of the primal solution by locating those points where the polynomial $q(t) = (F_{\text{lo}}^* u_{\text{est}})(t)$ has modulus equal to one. Once the support is estimated accurately, a solution to the primal problem can be found by solving a discrete problem. Figure 3 shows the result of applying this scheme to a simple example. We omit further details and defer the analysis of this approach to future work.




(a)




(b)

Figure 3: (a) Original support of a signal obeying the minimum-separation condition (black) along with its low-pass projection before (blue) and after adding noise (red). The low-pass projection is obtained by truncating the spectrum of the signal to its first 101 Fourier coefficients. The noise added to the noiseless Fourier coefficients is i.i.d. Gaussian with amplitude giving a signal-to-noise ratio of 31.5 dB. (b) Original signal (blue) and estimate obtained by solving the SDP (3.2) (red).

### Acknowledgements

E. C. is partially supported by AFOSR under grant FA9550-09-1-0643, by ONR under grant N00014-09-1-0258 and by a gift from the Broadcom Foundation. C. F. is supported by a Fundación Caja Madrid Fellowship. We thank Carlos Sing-Long for useful feedback about an earlier version of the manuscript.


## References


[1] B. N. Bhaskar, G. Tang, and B. Recht. Atomic norm denoising with applications to line spectral estimation. Preprint.

[2] G. Bienvenu. Influence of the spatial coherence of the background noise on high resolution passive methods. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '79.*, volume 4, pages 306 – 309, 1979.

[3] T. Blu, P. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot. Sparse sampling of signal innovations. *Signal Processing Magazine, IEEE*, 25(2):31–40, 2008.

[4] E. J. Candès and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics.* To appear.

[5] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank. Sensitivity to basis mismatch in compressed sensing. *IEEE Transactions on Signal Processing*, 59(5):2182–2195, 2011.

[6] D. L. Donoho. Superresolution via sparsity constraints. *SIAM J. Math. Anal.*, 23(5):1309–1331, 1992.

[7] M. F. Duarte and R. G. Baraniuk. Spectral compressive sensing. Preprint.

[8] B. Dumitrescu. *Positive Trigonometric Polynomials and Signal Processing Applications.* Springer, 2007.

[9] A. Fannjiang and W. Liao. Coherence-pattern guided compressive sensing with unresolved grids. *SIAM J. Imaging Sci.*, 5:179, 2012.

[10] H. Greenspan. Super-resolution in medical imaging. *Comput. J.*, 52:43–63, 2009.

[11] T. D. Harris, R. D. Grober, J. K. Trautman, and E. Betzig. Super-resolution imaging spectroscopy. *Appl. Spectrosc.*, 48(1):14A–21A, 1994.

[12] C. Helstrom. The detection and resolution of optical signals. *Information Theory, IEEE Transactions on*, 10(4):275 – 287, 1964.

[13] L. Hu, Z. Shi, J. Zhou, and Q. Fu. Compressed sensing of complex sinusoids: An approach based on dictionary refinement. *Signal Processing, IEEE Transactions on*, 60(7):3809 –3822, 2012.

[14] F. Itakura. Line spectrum representation of linear predictor coefficients of speech signals. *The Journal of the Acoustical Society of America*, 57(S1):S35–S35, 1975.

[15] J. Kennedy, O. Israel, A. Frenkel, R. Bar-Shalom, and H. Azhari. Super-resolution in PET imaging. *Medical Imaging, IEEE Transactions on*, 25(2):137 –147, 2006.

[16] V. Khaidukov, E. Landa, and T. J. Moser. Diffraction imaging by focusing-defocusing: An outlook on seismic superresolution. *Geophysics*, 69(6):1478–1490, 2004.

[17] B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *Annals of Statistics*, 28(5):1302–1053, 1992.

[18] J. Lindberg. Mathematical concepts of optical superresolution. *Journal of Optics*, 14(8):083001, 2012.

[19] D. Makovoz and F. R. Marleau. Point source extraction with MOPEX. *Publications of the Astronomical Society of the Pacific*, 117(836):1113–1128, 2005.

[20] C. W. McCutchen. Superresolution in microscopy and the Abbe resolution limit. *J. Opt. Soc. Am.*, 57(10):1190–1190, 1967.

[21] P. Milanfar, editor. *Super-Resolution Imaging.* Series: Digital Imaging and Computer Vision, 2010.

[22] J. Odendaal, E. Barnard, and C. Pistorius. Two-dimensional superresolution radar imaging using the MUSIC algorithm. *Antennas and Propagation, IEEE Transactions on*, 42(10):1386 –1391, 1994.

[23] S. C. Park, M. K. Park, and M. G. Kang. Super-resolution image reconstruction: a technical overview. *Signal Processing Magazine, IEEE*, 20(3):21 – 36, 2003.

[24] R. Prony. Essai expérimental et analytique. *J. de l'Ecole Polytechnique (Paris)*, 1(2):24–76, 1795.

[25] K. G. Puschmann and F. Kneer. On super-resolution in astronomical imaging. *Astronomy and Astrophysics*, 436:373–378, 2005.

[26] R. Roy and T. Kailath. ESPRIT- estimation of signal parameters via rotational invariance techniques. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 37(7):984 –995, 1989.

[27] W. Rudin. *Real and complex analysis.* McGraw-Hill Book Co., New York, 3rd edition, 1987.

[28] R. Schmidt. Multiple emitter location and signal parameter estimation. *Antennas and Propagation, IEEE Transactions on*, 34(3):276 – 280, 1986.

[29] M. Shahram and P. Milanfar. Imaging below the diffraction limit: a statistical analysis. *Image Processing, IEEE Transactions on*, 13(5):677 –689, 2004.

[30] M. Shahram and P. Milanfar. On the resolvability of sinusoids with nearby frequencies in the presence of noise. *Signal Processing, IEEE Transactions on*, 53(7):2579 – 2588, 2005.

[31] D. Slepian. Prolate spheroidal wave functions, Fourier analysis, and uncertainty. V - The discrete case. *Bell System Technical Journal*, 57:1371–1430, 1978.

[32] P. Stoica and N. Arye. MUSIC, maximum likelihood, and Cramér-Rao bound. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 37(5):720 –741, 1989.

[33] P. Stoica and P. Babu. Sparse estimation of spectral lines: Grid selection problems and their solutions. *Signal Processing, IEEE Transactions on*, 60(2):962 –967, 2012.

[34] P. Stoica, R. Moses, B. Friedlander, and T. Soderstrom. Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 37(3):378–392, 1989.

[35] P. Stoica and R. L. Moses. *Spectral Analysis of Signals.* Prentice Hall, 2005.

[36] P. Stoica and A. Nehorai. Statistical analysis of two nonlinear least-squares estimators of sine-wave parameters in the colored-noise case. *Circuits, Systems, and Signal Processing*, 8(1):3–15, 1989.

[37] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht. Compressed sensing off the grid. Preprint.

# A Proof of Lemma 2.5

We use the construction described in Section 2 of [4]. In more detail,

$$q(t) = \sum_{t_k \in T} \alpha_k G(t - t_k) + \beta_k G^{(1)}(t - t_k),$$

where $\alpha, \beta \in \mathbb{C}^{|T|}$ are coefficient vectors,

$$G(t) = \left[ \frac{\sin\left(\left(\frac{f_{\text{lo}}}{2} + 1\right)\pi t\right)}{\left(\frac{f_{\text{lo}}}{2} + 1\right)\sin(\pi t)} \right]^4, \quad t \in \mathbb{T} \setminus \{0\}, \tag{A.1}$$

and $G(0) = 1$; here, $G^{(\ell)}$ is the $\ell$th derivative of $G$. If $f_{\text{lo}}$ is even, $G(t)$ is the square of the Fejér kernel. By construction, the coefficients $\alpha$ and $\beta$ are selected such that for all $t_j \in T$,

$$q(t_j) = v_j$$
$$q'(t_j) = 0.$$

Without loss of generality we consider $t_j = 0$ and bound $q(t) - v_j$ in the interval $[0, 0.16\lambda_{\text{lo}}]$. To ease notation, we define $w(t) = q(t) - v_j = w_R(t) + i\, w_I(t)$, where $w_R$ is the real part of $w$ and $w_I$ the imaginary part. Leveraging different results from Section 2 in [4] (in particular equations (2.23) and (2.25) and Lemmas 2.2 and 2.7), we have

$$\begin{aligned}
|w_R''(t)| &= \left| \sum_{t_k \in T} \operatorname{Re}(\alpha_k)\, G^{(2)}(t - t_k) + \sum_{t_k \in T} \operatorname{Re}(\beta_k)\, G^{(3)}(t - t_k) \right| \\
&\leq ||\alpha||_{L_\infty} \sum_{t_k \in T} \left| G^{(2)}(t - t_k) \right| + ||\beta||_{L_\infty} \sum_{t_k \in T} \left| G^{(3)}(t - t_k) \right| \\
&\leq C_\alpha \left( \left| G^{(2)}(t) \right| + \sum_{t_k \in T \setminus \{0\}} \left| G^{(2)}(t - t_k) \right| \right) + C_\beta \lambda_{\text{lo}} \left( \left| G^{(3)}(t) \right| + \sum_{t_k \in T \setminus \{0\}} \left| G^{(3)}(t - t_k) \right| \right) \\
&\leq C\, f_{\text{lo}}^2.
\end{aligned}$$

The same bound holds for $w_I$. Since $w_R(0)$, $w_R'(0)$, $w_I(0)$ and $w_I'(0)$ are all equal to zero, this implies $|w_R(t)| \leq C' f_{\text{lo}}^2 t^2$ and $|w_I(t)| \leq C' f_{\text{lo}}^2 t^2$ in the interval of interest, which allows the conclusion

$$|w(t)| \leq C\, f_{\text{lo}}^2 t^2.$$

# B Proof of Lemma 2.7

The proof is similar to that of Lemma 2.4 (see Section 2 of [4]), where a low-frequency kernel and its derivative are used to interpolate an arbitrary sign pattern on a support satisfying the minimum-distance condition. More precisely, we set

$$q_1(t) = \sum_{t_k \in T} \alpha_k G(t - t_k) + \beta_k G^{(1)}(t - t_k), \tag{B.1}$$

where $\alpha, \beta \in \mathbb{C}^{|T|}$ are coefficient vectors, $G$ is defined by (A.1). Note that $G$, $G^{(1)}$ and, consequently, $q_1$ are trigonometric polynomials of degree at most $f_0$. By Lemma 2.7 in [4], it holds that for any $t_0 \in T$ and $t \in \mathbb{T}$ obeying $|t - t_0| \leq 0.16\lambda_{\text{lo}}$,

$$\sum_{t_k \in T \setminus \{t_0\}} \left| G^{(\ell)}(t - t_k) \right| \leq C_\ell f_{\text{lo}}^\ell, \tag{B.2}$$

where $C_\ell$ is a positive constant for $\ell = 0, 1, 2, 3$; in particular, $C_0 \leq 0.007$, $C_1 \leq 0.08$ and $C_2 \leq 1.06$. In addition, there exist other positive constants $C_0'$ and $C_1'$, such that for all $t_0 \in T$ and $t \in \mathbb{T}$ with $|t - t_0| \leq \Delta/2$,

$$\sum_{t_k \in T \setminus \{t_0\}} \left| G^{(\ell)}\left(t - t_k\right) \right| \leq C_\ell' f_{\text{lo}}^\ell \tag{B.3}$$

for $\ell = 0, 1$. We refer to Section 2.3 in [4] for a detailed description of how to compute these bounds.

In order to satisfy (2.18) and (2.19), we constrain $q_1$ as follows: for each $t_j \in T$,

$$\begin{aligned} q_1(t_j) &= 0, \\ q_1'(t_j) &= v_j. \end{aligned}$$

Intuitively, this forces $q_1$ to approximate the linear function $v_j\left(t - t_j\right)$ around $t_j$. These constraints can be expressed in matrix form,

$$\begin{bmatrix} D_0 & D_1 \\ D_1 & D_2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} 0 \\ v \end{bmatrix},$$

where

$$\left(D_0\right)_{jk} = G\left(t_j - t_k\right), \quad \left(D_1\right)_{jk} = G^{(1)}\left(t_j - t_k\right), \quad \left(D_2\right)_{jk} = G^{(2)}\left(t_j - t_k\right),$$

and $j$ and $k$ range from 1 to $|T|$. It is shown in Section 2.3.1 of [4] that under the minimum-separation condition this system is invertible, so that $\alpha$ and $\beta$ are well defined. These coefficient vectors can consequently be expressed as

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} -D_0^{-1} D_1 \\ I \end{bmatrix} S^{-1} v, \quad S := D_2 - D_1 D_0^{-1} D_1,$$

where $S$ is the Schur complement. Inequality (B.2) implies

$$\left|\left|I - D_0\right|\right|_\infty \leq C_0, \tag{B.4}$$

$$\left|\left|D_1\right|\right|_\infty \leq C_1 f_{\text{lo}}, \tag{B.5}$$

$$\left|\left|\kappa I - D_2\right|\right|_\infty \leq C_2 f_{\text{lo}}^2, \tag{B.6}$$

where $\kappa = \left|G^{(2)}(0)\right| = \pi^2 f_{\text{lo}}(f_{\text{lo}} + 4)/3$.

Let $\|M\|_\infty$ denote the usual infinity norm of a matrix $M$ defined as $\|M\|_\infty = \max_{\|x\|_\infty = 1} \|Mx\|_\infty = \max_i \sum_j |a_{ij}|$. Then, if $\left|\left|I - M\right|\right|_\infty < 1$, the series $M^{-1} = \left(I - (I - M)\right)^{-1} = \sum_{k \geq 0} \left(I - M\right)^k$ is convergent and we have

$$\left|\left|M^{-1}\right|\right|_\infty \leq \frac{1}{1 - \left|\left|I - M\right|\right|_\infty}.$$

This, together with (B.4), (B.5) and (B.6) implies

$$\begin{aligned} \left|\left|D_0^{-1}\right|\right|_\infty &\leq \frac{1}{1 - \left|\left|I - D_0\right|\right|_\infty} \leq \frac{1}{1 - C_0}, \\ \left|\left|\kappa I - S\right|\right|_\infty &\leq \left|\left|\kappa I - D_2\right|\right|_\infty + \left|\left|D_1\right|\right|_\infty \left|\left|D_0^{-1}\right|\right|_\infty \left|\left|D_1\right|\right|_\infty \leq \left(C_2 + \frac{C_1^2}{1 - C_0}\right) f_{\text{lo}}^2, \\ \left|\left|S^{-1}\right|\right|_\infty &= \kappa^{-1} \left|\left| \left(\frac{S}{\kappa}\right)^{-1} \right|\right|_\infty \leq \frac{1}{\kappa - \left|\left|\kappa I - S\right|\right|_\infty} \leq \left(\kappa - \left(C_2 + \frac{C_1^2}{1 - C_0}\right) f_{\text{lo}}^2\right)^{-1} \leq C_\kappa \lambda_{\text{lo}}^2, \end{aligned}$$

for a certain positive constant $C_\kappa$. Note that due to the numeric upper bounds on the constants in (B.2) $C_\kappa$ is indeed a positive constant as long as $f_{\text{lo}} \geq 1$. Finally, we obtain a bound on the magnitude of the entries of $\alpha$

$$\left|\left|\alpha\right|\right|_\infty = \left|\left|D_0^{-1} D_1 S^{-1} v\right|\right|_\infty \leq \left|\left|D_0^{-1} D_1 S^{-1}\right|\right|_\infty \leq \left|\left|D_0^{-1}\right|\right|_\infty \left|\left|D_1\right|\right|_\infty \left|\left|S^{-1}\right|\right|_\infty \leq C_\alpha \lambda_{\text{lo}}, \tag{B.7}$$

where $C_\alpha = C_\kappa C_1 / (1 - C_0)$, and on the entries of $\beta$

$$||\beta||_\infty = \left|\left|S^{-1} v\right|\right|_\infty \le \left|\left|S^{-1}\right|\right|_\infty \le C_\beta \lambda_{\mathrm{lo}}^2, \tag{B.8}$$

for a positive constant $C_\beta = C_\kappa$. Combining these inequalities with (B.3) and the fact that the absolute values of $G(t)$ and $G^{(1)}(t)$ are bounded by one and $7 f_{\mathrm{lo}}$ respectively (see the proof of Lemma C.5 in [4]), we have that for any $t$

$$\begin{aligned}
|q_1(t)| &= \left| \sum_{t_k \in T} \alpha_k G(t - t_k) + \sum_{t_k \in T} \beta_k G^{(1)}(t - t_k) \right| \\
&\le ||\alpha||_\infty \sum_{t_k \in T} |G(t - t_k)| + ||\beta||_\infty \sum_{t_k \in T} \left| G^{(1)}(t - t_k) \right| \\
&\le C_\alpha \lambda_{\mathrm{lo}} \left( |G(t)| + \sum_{t_k \in T \setminus \{t_i\}} |G(t - t_k)| \right) + C_\beta \lambda_{\mathrm{lo}}^2 \left( \left| G^{(1)}(t) \right| + \sum_{t_k \in T \setminus \{t_i\}} \left| G^{(1)}(t - t_k) \right| \right) \\
&\le C \lambda_{\mathrm{lo}},
\end{aligned} \tag{B.9}$$

where $t_i$ denotes the element in $T$ nearest to $t$ (note that all other elements are at least $\Delta/2$ away). Thus, (2.19) holds.

The proof is completed by the following lemma, which proves (2.18).

**Lemma B.1** *For any $t_j \in T$ and $t \in \mathbb{T}$ obeying $|t - t_j| \le 0.16 \lambda_{\mathrm{lo}}$, we have*

$$|q_1(t) - v_j (t - t_j)| \le \frac{C (t - t_j)^2}{\lambda_{\mathrm{lo}}}.$$

**Proof** We assume without loss of generality that $t_j = 0$. By symmetry, it suffices to show the claim for $t \in (0, 0.16\, \lambda_{\mathrm{lo}}]$. To ease notation, we define $w(t) = v_j t - q_1(t) = w_R(t) + i\, w_I(t)$, where $w_R$ is the real part of $w$ and $w_I$ the imaginary part. Leveraging (B.7), (B.8) and (B.2) together with the fact that $G^{(2)}(t)$ and $G^{(3)}(t)$ are bounded by $4 f_{\mathrm{lo}}^2$ and $6 f_{\mathrm{lo}}^3$ respectively if $|t| \le 0.16 \lambda_{\mathrm{lo}}$ (see the proof of Lemma 2.3 in [4]), we obtain

$$\begin{aligned}
|w_R''(t)| &= \left| \sum_{t_k \in T} \mathrm{Re}(\alpha_k)\, G^{(2)}(t - t_k) + \sum_{t_k \in T} \mathrm{Re}(\beta_k)\, G^{(3)}(t - t_k) \right| \\
&\le ||\alpha||_\infty \sum_{t_k \in T} \left| G^{(2)}(t - t_k) \right| + ||\beta||_\infty \sum_{t_k \in T} \left| G^{(3)}(t - t_k) \right| \\
&\le C_\alpha \lambda_{\mathrm{lo}} \left( \left| G^{(2)}(t) \right| + \sum_{t_k \in T \setminus \{0\}} \left| G^{(2)}(t - t_k) \right| \right) + C_\beta \lambda_{\mathrm{lo}}^2 \left( \left| G^{(3)}(t) \right| + \sum_{t_k \in T \setminus \{0\}} \left| G^{(3)}(t - t_k) \right| \right) \\
&\le C\, f_{\mathrm{lo}}.
\end{aligned}$$

The same bound applies to $w_I$. Since $w_R(0)$, $w_R'(0)$, $w_I(0)$ and $w_I'(0)$ are all equal to zero, this implies $|w_R(t)| \le C f_{\mathrm{lo}} t^2$—and similarly for $|w_I(t)|$—in the interval of interest. Whence, $|w(t)| \le C f_{\mathrm{lo}} t^2$. ■

# C Extension to multiple dimensions

Lemmas 2.4 (together with Lemma 2.5) and 2.7 construct bounded low-frequency polynomials which interpolate a sign pattern on a well-separated set of points $S$ and have bounded second derivatives in a neighborhood

of $S$. In order to extend our results all we need is to prove their multidimensional analogs (in this case, instead of bounding the second derivative, we must bound the eigenvalues of the Hessian matrix). One can proceed in a way similar to the proof of Lemmas 2.4 and 2.7, namely, by using a low-frequency kernel constructed by tensorizing several squared Fejér kernels to interpolate the sign pattern, while constraining the first-order derivatives to either vanish or have a fixed value. To do this, we can set up a system of equations and prove that it is well conditioned using the rapid decay of the interpolation kernel away from the origin. Finally, one can verify that the construction satisfies the required conditions by exploiting the fact that the interpolation kernel and its derivatives are locally quadratic and rapidly decaying. This is spelled out in the proof of Proposition C.1 in [4] to prove a version of Lemma 2.4 in two dimensions.